An Ontological Solution to the Sleeping Beauty Problem

Paul Franceschi
University of Corsica

p.franceschi@univ-corse.fr
http://www.univ-corse.fr/~franceschi

ABSTRACT. I describe in this paper an ontological solution to the Sleeping Beauty problem. I begin with describing the Entanglement urn experiment. I restate first the Sleeping Beauty problem from a wider perspective than the usual opposition between halfers and thirders. I also argue that the Sleeping Beauty experiment is best modelled with the Entanglement urn. I draw then the consequences of considering that some balls in the Entanglement urn have ontologically different properties form normal ones. The upshot is that I endorse the halfer conclusion on the probability of Heads once beauty is awaken and the thirder conclusion on conditional probabilities, and that original conclusions ensue on the probability of waking on Monday.

1. The Entanglement urn

Let us consider the following experiment. In front of you is a urn. The experimenter asks you to study very carefully the properties of the balls that are in the urn. You go up then to the urn and begin to examine carefully its content. You note first that the urn contains only red or green balls. By curiosity, you decide to take a sample in the urn of a red ball. Surprisingly, you notice that while you catch this red ball, another ball, but of green colour, also moves simultaneously. You decide then to replace the red ball in the urn and you notice that immediately, the latter green ball also springs back in the urn. Intrigued, you decide then to catch this green ball. You note then that the red ball also goes out of the urn at the same time. Furthermore, while you replace the green ball in the urn, the red ball also springs back at the same time at its initial position in the urn. You decide then to withdraw another red ball from the urn. But while it goes out of the urn, nothing else occurs. Taken aback, you decide then to undertake a systematic and rigorous study of all balls present in the urn.

At the end of several hours of a meticulous examination, you are now capable of describing precisely the properties of the balls present in the urn. The latter contains in total 1000 red balls and 500 green balls. Among the red balls, 500 are completely normal balls. But 500 other red balls have completely surprising properties. Indeed, each of them is linked to a different green ball. When you take away one of these red balls, the green ball which is linked to it also goes out at the same time of the urn, as though it was linked to the red ball by a magnetic force. The red ball and the green ball which is linked to it behave then as one single object. Indeed, if you take away the red ball from the urn, the linked green ball is also extracted instantly. And conversely, if you withdraw from the urn one of the green balls, the red ball which is linked to it goes out immediately of the urn. You even tried to destroy one of the balls of a linked pair of balls, and you noticed that in such case, the ball of the other colour which is indissociably linked to it was also destroyed instantaneously. Indeed, it appears to you that these pairs of balls behave as one single object.

The functioning of this urn leaves you somewhat perplexed. In particular, your are intrigued by the properties of the pairs of correlated balls. After reflection, you tell yourself that the properties of the pairs of correlated balls are finally in all respects identical to those of two entangled quantum objects. The entanglement (Aspect & al. 1982) is indeed the phenomenon which links up two photons, for example, so that when one modifies the quantum state of one of the entangled photons, the quantum state of the other one is instantly modified accordingly, whatever the distance where it is situated. Indeed, the pair of entangled photons really behave as one and the same object. You decide to call "Entanglement urn" this urn with its astonishing properties. After reflection, what appears peculiar in this urn, is that it includes at the same time normal and entangled balls. The normal red balls have nothing different with our familiar balls. But entangled balls behave in a completely different way. What is amazing, you think, is that nothing seemingly differentiates the normal red balls from the red entangled ones. You tell yourself finally that it could be confusing.

Your reflection on the pairs of entangled balls and their properties also leads you to question the way the balls which compose the pairs of entangled balls are to be counted. Are they counted as normal balls? Or do specific rules govern the way pairs of entangled balls are counted? You add a normal red ball in an Entanglement urn. It is then necessary to increment the number of red balls present in the urn. On the other hand, the total number of green balls is unaffected. But what is it when you add in the Entanglement urn the red ball of a pair of entangled balls? In that case, the linked green ball of the same pair of entangled balls is also added in the urn instantly. Hence, when you add a red ball of a pair of entangled balls in the urn, it proves to be that you also add at the same time, its linked green ball. So, in that case, you must increment the total number of red balls, but also the total number of green balls present in the urn. In the same way, if you subtract a normal red ball from the urn, you simply decrement the total number of red balls of the urn, without changing the number of green balls present in the urn. But if you remove a red ball (resp. green) of a pair of entangled balls, you must decrement the total number of red balls (resp. green) present in the urn as well as the total number of green balls (resp. red).

At this very moment, the experimenter happens again and withdraws all balls from the urn. He announces that you are going to participate in the following experiment:

> *The Entanglement urn* A fair coin will be randomly tossed. If the coin lands Heads, the experimenter will put a normal red ball in the urn. On the other hand, if the coin lands Tails, he will put a pair of entangled balls in the urn, composed of a red ball and a green ball, both indissociably linked. The experimenter also adds that the room will be put in absolute darkness, and that you will therefore be completely unable to detect the colour of the balls, no more that you will be able to know, when you will have withdrawn a ball from the urn, whether it is a normal ball, or a ball which is part of a pair of entangled balls. The experimenter tosses the coin and while you catch a ball from the urn, he asks you to assess the likelihood that the coin felt Heads.

2. The Sleeping Beauty problem

Consider now the well-known *Sleeping Beauty problem* (Elga 2000, Lewis 2001). Sleeping Beauty learns that she will be put into sleep on Sunday by some researchers. A fair coin will be tossed and if the coin lands Heads, Beauty will be awaken once on Monday. On the other hand, if the coin lands Tails, Beauty will be awaken twice: on Monday and on Tuesday. After each waking, she will be put into sleep again and she will forget that waking. On awakening on Monday, what should be Beauty's credence that the coin did land Heads?

At this step, one obvious first answer (I) goes as follows: since the coin is fair, the initial probability that the coin lands Head is 1/2. And during the course of the experiment, Sleeping Beauty doesn't get any novel information. Hence, the probability of Heads still remains 1/2.

By contrast, an alternative reasoning (II) runs as follows. Suppose that the experiment is repeated many times, say, to fix ideas, 1000 times. Then there will be approximately 500 Heads-wakings on Monday, 500 Tails-wakings on Monday and 500 Tails-wakings on Tuesday. Hence, the reasoning goes, the probability of Heads equals 500/1500 = 1/3.

The argument for 1/2 and the argument for 1/3 yield conflicting conclusions. But these two concurrent lines of reasoning are also accompanied with a calculation of the probability of waking on Monday and on Tuesday. To simplify matters, a Monday waking can be modelled with a red ball, and a Tuesday waking with a green ball. Now from the halfer perspective, the probability P(R) of drawing a red ball (this is tantamount to the probability of waking on Monday) is such that P(R) = 3/4. And conversely, the probability P(G) of drawing a green ball (this is tantamount to the probability of waking on Tuesday) is such that P(G) = 1/4. By contrast, from the thirder's perspective, the probability P(R) of drawing a red ball equals 2/3 and the probability P(G) of drawing a green ball equals 1/3.

But this is not the whole story. In effect, the argument for 1/2 and for 1/3 also have their own account of conditional probabilities. To begin with, the probability P(Heads|G) of Heads on drawing a green ball is not a subject of disagreement, for it equals 0 in both accounts. The same goes for the probability P(Tails|G) of Tails on drawing a green ball, since it equals 1 from the halfer's or from the thirder's viewpoint. But agreement stops when one considers the probability P(Heads|R) of Heads on drawing a red ball. For P(Heads|R) = 2/3 for a halfer and P(Heads|R) = 1/2 from a thirder's perspective. On the other hand, the probability P(Tails|R) of Tails on drawing a red ball is 1/3 from a halfer standpoint, and 1/2 for a thirder.

3. An ontological account based on the Entanglement urn

In what follows, I shall present an ontological solution to the Sleeping Beauty problem, which rests basically on the Entanglement urn experiment. A consequence of this account is that it incorporates insights from the halfer and thirder standpoints, a line of resolution initiated by Nick Bostrom (2007)[1]. The Sleeping Beauty problem is usually presented as a problem arising from conflicting conclusions resulting from two concurrent lines of assigning the probability of Heads once Beauty is awaken: the argument for 1/2 and the argument for 1/3. I shall argue, however, that this description of the Sleeping Beauty problem is misleading and that we need to envisage the issue from a wider perspective. For present purposes, the Sleeping Beauty problem is the issue of calculating properly (i) the probability of Heads (resp. Tails) once Beauty is awaken; (ii) the probability of a waking on Monday (resp. Tuesday); and (iii) the probability of Heads (resp. Tails) on waking on Monday. In this broader context, I shall argue that the halfer's response is right with regard to the first question, that the thirder's answer is right with respect to the second question, and that both halfer's and thirder's responses are wrong with regard to the third question. The reminder of this paper will explain how these prima facie surprising results are consistently obtained.

To begin with, the present solution will endorse the halfer conclusion, with regard to the probability of Heads, once Beauty is awaken. I shall argue then that the argument for 1/3 is fallacious and that the *Entanglement urn* experiment casts light on the flaw in reasoning (II). Let us examine this in more detail. To begin with, it appears that the argument for 1/3 is based on a urn analogy. This analogy associates the situation inherent to the Sleeping Beauty experiment with a urn that contains, in the long run (assuming that the experiment is repeated 1000 times), 500 red balls (Heads-wakings on Monday), 500 red balls (Tails-wakings on Monday) and 500 green balls (Tails-wakings on Tuesday), i.e. 1000 red balls and 500 green balls. The corresponding urn contains then 1000 red balls and 500 green balls. For present purposes, let us call this sort of urn a “standard urn”. Now it appears that the argument for 1/3 is based on a urn analogy with this standard urn and that the probability of Heads is determined by the ratio of the number of Heads-wakings among the total number of wakings. At this step, a question arises: is the analogy with the *standard urn* well-grounded in the argument for 1/3? In other terms, isn't another urn model best suited to the iterated Sleeping Beauty experiment? In the present context, this alternative can be formulated more accurately as follows: isn't the situation inherent to the Sleeping Beauty experiment better put in analogy with the *Entanglement urn*, rather than with the *standard urn*? It strikes me that the standard urn fails to do justice to an essential feature of the experiment: in the Tails case, the Monday waking and the Tuesday waking are correlated; while on the other hand, the Monday waking in the Heads case is independent. Let us examine this in more detail.

The intuition underlying the argument for 1/3 in the Sleeping Beauty experiment is that one is entitled to add unrestrictedly red and green balls to compute frequencies. However, I shall argue that this intuition is misleading, as the Entanglement urn experiment suggests. For in reasoning (II), one feels intuitively entitled to add red-Heads (Heads-wakings on Monday), red-Tails (Tails-wakings on Monday) and green-Tails (Tails-wakings on Tuesday) balls to compute frequencies. But red-Heads and red-Tails balls appear to be objects of a fundamentally different nature in the present context. In effect, red-Heads balls are in all respects similar to our familiar objects, and can be considered properly as single objects. By contrast, it appears that red-Tails balls are quite indissociable from green-Tails balls. For we cannot draw a red-Tails ball without picking the associated green-Tails ball. And conversely, we cannot draw a green-Tails ball without picking the associated red-Tails ball. In this sense, red-Tails balls and the associated green-Tails balls do not behave as our familiar objects, but are much similar to entangled quantum objects. For Monday-Tails wakings are indissociable from

[1] Bostrom opens the path to a third way out to the Sleeping Beauty problem: “At any rate, one might hope that having a third contender for how Beauty should reason will help stimulate new ideas in the study of self-location”. In his account, Bostrom sides with the halfer on P(Heads) and with the thirder on conditional probabilities, but his treatment has some counter-intuitive consequences on conditional probabilities. As far as I can see, the present account is devoid of these drawbacks on conditional probabilities.

Tuesday-Tails wakings. And Beauty cannot be awaken on Monday (resp. Tuesday) without being awaken on Tuesday (resp. Monday). From this viewpoint, it is mistaken to consider red-Tails and green-Tails balls as separate objects. The correct intuition is that the a red-Tails and the associated green-Tails ball are a pair of entangled balls and constitute but one single object. In this context, red-Tails and green-Tails balls are best seen intuitively as constituents and mere parts of one single object. In other words, red-Heads balls and, on the other hand, red-Tails and green-Tails balls, cannot be considered as objects of the same type for probability purposes. And this situation justifies the fact that one is not entitled to add unrestrictedly red-Heads, red-Tails and green-Tails balls to compute probability frequencies. For in this case, one adds objects of intrinsically different types, i.e. one single object with the mere part of another single object.

Now the key point appears to be the following one. Consider the Entanglement urn. Normal red balls behave as usual. But entangled balls do behave differently, with regard to statistics. Suppose I add the red ball of an entangled pair in the Entanglement urn. Then I also add instantly in the urn the associated green ball of the entangled pair. Suppose, conversely, that I remove the red ball of an entangled pair from the urn. Then I also remove instantly the associated green ball. The same goes now for Sleeping Beauty, as the analogy suggests. And the consequences are not so that innocuous. What is the probability of a waking on Monday? This is tantamount to calculating the probability P(R) of drawing a red ball in the Entanglement urn? On Heads, the probability of drawing a red ball is 1. On Tails, we can either draw the red or the green ball of an entangled pair. But it should be pointed out that if we pick on Tails the green ball of an entangled pair, we also draw instantly the associated red ball. Hence, the probability of drawing a red ball on Tails is also 1. Thus, P(R) = 1/2 x 1 + 1/2 x 1 = 1. Conversely, what is the probability of a waking on Tuesday? This is tantamount to the probability P(G) of drawing a green ball. The probability of drawing a green ball is 0 in the Heads case, but 1 in the Tails case. For in the latter case, either we draw the green or the red ball of an entangled pair. But even if we draw the red ball of the entangled pair, we draw then instantly the associated green ball. Hence, P(G) = 1/2 x 0 + 1/2 x 1 = 1/2. To sum up: P(R) = 1 and P(G) = 1/2. The probability of a waking on Monday is then 1, and the probability of a waking on Tuesday is 1/2.

From the above, it results that P(R) + P(G) = 1 + 1/2 = 1,5. In the present account, this results from the fact that drawing a red ball and drawing a green ball – in general – are *not exclusive events*. And – in particular – drawing a red ball and drawing a green ball from an entangled pair are not exclusive events for probability purposes. For we cannot draw the a red-Tails (resp. green-Tails) ball without drawing the associated green-Tails (resp. red-Tails) ball. The fact that drawing a red-Tails and drawing a green-Tails ball are not exclusive events is overlooked in the argument for 1/3, and notably in Elga's (Elga 2000) formulation of the argument for 1/3. For Elga enumerates first three possibilities: drawing a red-Heads ball (in Elga's terminology, H1: Heads and it is Monday), a red-Tails ball (T1: Tails and it is Monday) or a green-Tails ball (T2: Tails and it is Tuesday). Elga infers then that the probability of these three events are equal: P(H1) = P(T1) = P(T2). And, the reasoning goes, as these probabilities sum to 1, P(H1) = P(T1) = P(T2) = 1/3. But as we did see it, the flaw in this reasoning now appears clearly: it is the step that considers that these three probabilities sum to 1. In effect, as we did see it, these three events are not exclusive of one another, causing their overall probabilities sum to 1,5.

Now the preceding developments also have some consequences on conditional probabilities. Let us recall first how these latter are calculated on the two concurrent standard lines of reasoning. To begin with, the probability P(Heads|G) of Heads on drawing a green ball is not a subject of disagreement for halfers and thirders, since it equals 0 in both accounts. The same goes for the probability P(Tails|G) of Tails on drawing a green ball, since it equals 1 from a halfer or thirder viewpoint. But agreement stops when one considers the probability P(Heads|R) of Heads on drawing a red ball. For P(Heads|R) = 2/3 for a halfer and P(Heads|R) = 1/2 from a thirder's perspective. On the other hand, the probability P(Tails|R) of Tails on drawing a red ball is 1/3 from a halfer standpoint, and 1/2 for a thirder. In the present account, P(Heads|G) = 0 and P(Tails|G) = 1, as usual. But P(Heads|R) is calculated as follows. P(Heads|R) = [P(Heads) x P(R|Heads)] / P(R) = [1/2 x 1] / 1 = 1/2. And the same goes for P(Tails|R): P(Tails|R) = [P(Tails) x P(R|Tails)] / P(R) = [1/2 x 1] / 1 = 1/2. The upshot is that conditional probabilities are calculated in the same way as from a thirder's perspective.

Finally, the above results are summarised in the following table:

	halfer	*thirder*	*present account*
P(Heads)	**1/2**	1/3	**1/2**
P(Tails)	**1/2**	2/3	**1/2**
P(waking on Monday) ≡ P(R)	3/4	2/3	**1**
P(waking on Tuesday) ≡ P(G)	1/4	1/3	**1/2**
P(Heads| waking on Monday) ≡ P(Heads|R)	2/3	**1/2**	**1/2**
P(Tails| waking on Monday) ≡ P(Tails|R)	1/3	**1/2**	**1/2**

4. Handling the variations of the Sleeping Beauty problem

From the above, it follows that the present treatment of the Sleeping Beauty problem, is capable of handling several variations of the original problem which have recently flourished in the literature. For the above solution to the Sleeping Beauty problem applies straightforwardly, I shall argue, to these variations of the original experiment. Let us consider, to begin with, a variation were on Heads, Sleeping Beauty is not awaken on Monday but instead on Tuesday. This is modelled with an Entanglement urn that receives one normal green ball (instead of a red one in the original experiment) in the Heads case.

Let us suppose, second, that Sleeping Beauty is awaken two times on Monday in the Tails case (instead of being awaken on both Monday and Tuesday). This is then modelled with an Entanglement urn that receives one pair of entangled balls which are composed of two red balls in the Tails case. (instead of a pair of entangled balls composed of a red and a green ball in the original experiment).

Let us imagine, third, that Beauty is awaken two times – on Monday and Tuesday – in the Heads case, and three times – on Monday, Tuesday and Wednesday – in the Tails case. This is then modelled with an Entanglement urn that receives one pair of entangled balls composed of one red ball and one green ball in the Heads case; in the Tails case, the Entanglement urn is filled with one triplet of entangled balls, composed of a red, a green and a blue ball.

Let us also envisage another variation:[2] on Heads, Beauty is awaken once either on Monday or on Tuesday (determined randomly). In this variant, from the halfer's viewpoint, we get: P(Heads = 1/2); in addition, P(R) = P(G) = 1/2; and finally, P(Heads|R) = 1/2. On the other hand, P(Heads) = 1/3 from the thirder's perspective; and P(R) = P(G) = 1/2; lastly, P(Heads|R) = 1/3. In the present account, this is modelled with a urn that contains one red ball or one green ball in the Heads case, and one pair of red and green entangled balls. We get then accordingly: P(Heads) = 1/2, P(R) = 1/2 x 1/2 + 1 x 1/2 = 3/4 and P(G) = 1/2 x 1/2 + 1 x 1/2 = 3/4. And also: P(Heads|R) = 1/2.

Finally, the lesson of the Sleeping Beauty Problem appear to be the following: our current and familiar objects or concepts such as balls, wakings, etc. should not be considered as the sole relevant classes of objects for probability purposes. We should bear in mind that according to an unformalised axiom of probability theory, a given situation is standardly modelled with the help of urns, dices, balls, etc. But the rules that allow for these simplifications lack an explicit formulation. However in certain situations, in order to reason properly, it is also necessary to take into account somewhat unfamiliar objects whose constituents are pairs of indissociable balls or of mutually inseparable wakings, etc. This lesson was anticipated by Nelson Goodman, who pointed out in *Ways of Worldmaking* that some objects which are prima facie completely different from our familiar objects also deserve consideration: ‘we do not welcome molecules or concreta as elements of our everyday world, or combine tomatoes and triangles and typewriters and tyrants and tornadoes into a single kind’.[3] As we did see it, in some cases, we cannot add unrestrictedly an object of the Heads-world with an object of the Tails-world. For drawing a red ball and drawing a green ball in the Tails-world are not exclusive events. The upshot is that, in order to preserve our standard way of evaluating probabilities, we need to incorporate the fact that drawing one ball of an entangled pair is not exclusive of drawing the associated ball. And the status of our paradigm probabilistic object, namely a ball, appears to be

[2] I thank Laurent Delabre for pointing out this variation to me (personal correspondence).
[3]Goodman (1978, p. 21).

world-relative, since it can be a whole in the Heads-world and a part in the Tails-world. Once this goodmanian step accomplished, we should be less vulnerable to certain subtle cognitive traps in probabilistic reasoning.[4]

References

Arntzenius, F. (2002) Reflections on Sleeping Beauty, *Analysis*, 62-1, 53-62

Aspect, A., Dalibard, J. & Roger, G. (1982) Physical Review Letters. 49, 1804-1807.

Bostrom, N. (2002) *Anthropic Bias: Observation Selection Effects in Science and Philosophy,* New York: Routledge

Bostrom, N. (2007) Sleeping Beauty and Self-Location : A Hybrid Model, *Synthese*, 157, 59-78

Bradley, D. (2003) Sleeping Beauty: a note on Dorr's argument for 1/3, *Analysis*, 63, 266-268

Delabre, L. (2008) La Belle au Bois dormant: débat autour d'un paradoxe, manuscript.

Dorr, C. (2002) Sleeping Beauty: in Defence of Elga, *Analysis*, 62, 292-296

Elga, A. (2000) Self-locating Belief and the Sleeping Beauty Problem, *Analysis*, 60, 143-147

Goodman, N. (1978) *Ways of Worldmaking*, Indianapolis: Hackett Publishing Company

Groisman, B. (2008) The End of Sleeping Beauty's Nightmare, British Journal for the Philosophy of Science, 59(3):409-416

Lewis, D. (1983) Attitudes de dicto and de se, in *Philosophical Papers*, volume I, 133-159, New York: Oxford University Press

Lewis, D. (2001) Sleeping Beauty: Reply to Elga, *Analysis*, 61, 171-176

Monton, B. (2002) Sleeping Beauty and the Forgetful Bayesian, *Analysis*, 62, 47-53

White, R. (2006) The generalized Sleeping Beauty problem : A challenge for thirders, *Analysis*, 66, 114-119.

[4] I thank Jean-Paul Delahaye and Claude Panaccio for useful discussion. Special thanks are due to Laurent Delabre for stimulating correspondence.